

What is claimed is:

1        1. A method for selecting segments from a corpus of source utterances for  
2 synthesizing a target utterance, comprising:

3                searching a graph in which each path through the graph identifies a sequence of  
4 segments of the source utterances and a corresponding sequence of unit labels that  
5 characterizes a pronunciation of a concatenation of that sequence of segments, each path  
6 being associated with a numerical score that characterizes a quality of the sequence of  
7 segment;

8                wherein searching the graph includes matching a pronunciation of the target  
9 utterance to paths through the graph, and selecting segments for synthesizing the target  
10 utterance based on numerical scores of matching paths through the graph.

1        2. The method of claim 1 wherein selecting segments for synthesizing the  
2 target utterance includes identifying a path through the graph that matches the  
3 pronunciation of the target utterance and selecting the sequence of segments that is  
4 identified by the determined path.

1        3. The method of claim 2 wherein determining the path includes determining  
2 a best scoring path through the graph.

1        4. The method of claim 3 wherein determining the best scoring path involves  
2 using a dynamic programming algorithm.

1        5. The method of claim 2 further comprising concatenating the selected  
2 sequence of segments to form a waveform representation of the target utterance.

1        6. The method of claim 1 wherein selecting the segments for synthesizing the  
2 target utterance includes determining a plurality of paths through the graph that each  
3 matches the representation of the pronunciation of the target utterance.

1           7.     The method of claim 6 wherein selecting the segments further includes  
2 forming a plurality of sequences of segments, each associated with a different one of the  
3 plurality of paths.

1           8.     The method of claim 7 wherein selecting the segments further includes  
2 selecting one of the sequences of segments based on characteristics of those sequences of  
3 segments not determined by the corresponding sequences of unit labels associated with  
4 those sequences.

1           9.     The method of claim 1 further comprising forming a representation of a  
2 plurality of pronunciations of the target utterance, and wherein searching the graph  
3 includes matching any of the pronunciations of the target utterance to paths through the  
4 graph.

1           10.    The method of claim 1 further comprising forming a representation of the  
2 pronunciation of the target utterance in terms of alternating unit labels and transitions  
3 labels.

1           11.    The method of claim 1 wherein the graph includes a first part that encodes  
2 a sequence of segments and a corresponding sequence of unit labels for each of the  
3 source utterances, and a second part that encodes allowable transitions between segments  
4 of different source utterances and encodes a transition score for each of those transitions;  
5 and

6           matching the pronunciation of the target utterance to paths through the graph  
7 includes considering paths in which each transition between segments of different source  
8 utterances identified by that path corresponds to a different subpath of that path that  
9 passes through the second part of the graph.

1           12.    The method of claim 10, wherein selecting the segments for synthesis  
2 includes evaluating a score for each of the considered paths that is based on the transition  
3 scores associated with the subpaths through the second part of the graph.

1        13.     The method of claim 10 wherein a size of the second part of the graph is  
2 substantially independent of a size of the source corpus, and a complexity of matching  
3 the pronunciation through the graph grows less than linearly with the size of the corpus.

1        14.     The method of claim 1 further comprising:

2              providing the corpus of source utterances, each source utterance being segmented  
3 into a sequence of segments, each consecutive pair of segments in a source utterance  
4 forming a segment boundary, and each speech segment being associated with a unit label  
5 and each segment boundary being associated with a transition label; and

6              forming the graph, including forming a first part of the graph that encodes a  
7 sequence of segments and a corresponding sequence of unit labels for each of the source  
8 utterances, and forming a second part that encodes allowable transitions between  
9 segments of different source utterances and encodes a transition score for each of those  
10 transitions.

1        15.     The method of claim 14 wherein forming the second part of the graph is  
2 performed independently of the utterances in the corpus of source utterances.

1        16.     The method of claim 14 further comprising:

2              augmenting the corpus of source utterances with additional utterances; and

3              augmenting the graph including augmenting the first part of the graph to encode  
4 the additional utterances, and linking the augmented first part to the second part without  
5 modifying the second part based on the additional utterances.

1        17.     The method of claim 1 wherein the graph is associated with a finite-state  
2 transducer which accepts input symbols that include unit labels and transition labels, and  
3 that produces identifiers of segments of the source utterances, and wherein searching the  
4 graph is equivalent to composing a finite-state transducer representation of a  
5 pronunciation of the target utterance with the finite-state transducer with which the graph  
6 is associated.

1        18. Software stored on a computer-readable medium for causing a computer to  
2 perform functions comprising selecting segments from a corpus of source utterances for  
3 synthesizing a target utterance, wherein selecting the segments comprises:

4            searching a graph in which each path through the graph identifies a sequence of  
5 segments of the source utterances and a corresponding sequence of unit labels that  
6 characterizes a pronunciation of a concatenation of that sequence of segments, each path  
7 being associated with a numerical score that characterizes a quality of the sequence of  
8 segment;

9            wherein searching the graph includes matching a pronunciation of the target  
10 utterance to paths through the graph, and selecting segments for synthesizing the target  
11 utterance based on numerical scores of matching paths through the graph.